JERICO-S3 DELIVERABLE

Joint European Research Infrastructure for Coastal Observatories
Science, Services, Sustainability

DELIVERABLE #, WP# and full title	JERICO-S3 D.7.8 - WP7 - "Intelligent services and data science methodologies for the JIIM and the VA einfrastructure"
5 Key words	Automated Marine Observing Systems; Coastal Environment; Artificial Intelligence
Lead beneficiary	CNR
Lead Author	Simone Marini (CNR)
Co-authors	Eric Delory, Jerome Blandin, Alain Lefebvre, Enoc Martinez
Contributors	Bertrand Moreau, Joaquin Del Rio Fernandez, Emilie Poisson Caillault, Andrés Cianca, Dominique Durand, Laurent Delauney
Final version date/ Submission date	30-01-2023 – 13-04-2023

Nature:R
(R = Report, P = Prototype, D = Demonstrator, O = Other)

Dissemination level : **PU**
PU = Public, PP = Restricted to other programme participants (including the Commission Services), RE = Restricted to a group specified by the consortium (including the Commission Services), CO = Confidential, only for members of the consortium (including the Commission Services)

GRANT N°: 871153
PROJECT ACRONYME : JERICO-S3
PROJECT NAME : Joint European Research Infrastructure for Coastal Observatories - Science, services, sustainability
COORDINATOR : Laurent DELAUNEY - Ifremer, France - jerico@ifremer.fr

DOCUMENT TECHNICAL DESCRIPTION

Document ID	JERICO-S3-WP7-D7.8-30.01.2023-V1.0

REVISION HISTORY

Revision	Date	Modification	Author
V1.0	2023/01/10	First Draft	Simone Marini
V1.1	2023/01/30	Final version	Simone Marini

APPROVALS

	Name	Organisation	Date	Visa
Coordinator	Delauney Laurent	Ifremer	13-04-2023	X
WP Leaders	Simone Marini Eric Delory	CNR PLOCAN	30-01-2023	X

Diffusion list

Consortium beneficiaries	Third parties	Associated Partners	other
X	Univ-Litoral		

TABLE OF CONTENT

1. EXECUTIVE SUMMARY

This report describes the enabling data science methodologies for the innovative observation services and systems developed within the JERICO-S3 project. The information presented in this report enables the interpretation and modeling of relevant biological and ecological processes based on the data produced by the innovative sensors experimented within the JERICO-S3 project.

After the introduction of the main targets of the technology innovation activities within the JERICO-S3 project, the report provides in section 3.1 a short description of the coastal EMSO Generic Instrument Module (cEGIM), focusing on the high level characteristics that allow the execution of the intelligent services. Then, in section 3.2 the data processing capabilities of the JERICO-S3 partners in the analysis of biological, physical and ecological data are presented. In section 3.3 the intelligent services are specified comprising the data processing algorithm, the logic enabling the adaptive data sampling, based on the output of the data processing algorithm and the data transfer and visualization pipeline. Section 3.4 summarizes the dissemination activity related to the WP7 outcome.

2. INTRODUCTION

The technological developments of the JERICO-S3 EU project aims to strengthen and expand the infrastructure of the European network of coastal observatories. This objective is achieved with new observing systems and platforms equipped with new technologies for interoperability, innovative sensor packages for multidisciplinary ecosystem monitoring, coupling physics, biogeochemistry and biology. The planned technological developments consist in adapting interoperability standards, inter alia from the NeXOS and EMSODev European projects, developing on-board and on-server smart solutions for adaptive sampling, integrating technologies into dedicated sensor packages, further developing a capacity for high-frequency measurement of low trophic-level biological diversity and contaminants; hence filling critical gaps in the observation of the coastal ocean.

The task T7.4 on enabling data science for innovative observation services and systems, defines and develops data science methodologies for the interpretation and modeling of relevant biological and ecological processes based on the data produced by the sensors experimented within the JERICO-S3 project.

The data science methodologies implemented within the T7.4 activities are used to define and implement the intelligent services for the coastal EGIM module (cEGIM). Such intelligent services are capable to trigger cEGIM sensors and sampler, and adapt sensor configurations (e.g. data sampling frequency) for the automated adaptation of the cEGIM behavior to the dynamic relevant environmental conditions.

The data processing algorithms developed within the JERICO-S3 project will be also integrated into the JERICO-S3 e-infrastructure as part of the virtual catalog developed in the T7.5 activities.

3.MAIN REPORT

3.1.The Coastal EMSO Generic Instrument Module (cEGIM)

The EGIM (EMSO Generic Instrument Module) concept was developed in the EMSOdev project to meet the requirement of measuring consistently and continuously the same set of Essential Ocean Variables (EOVs) in a number of open ocean locations (Lantéri et al., 2019). Building on this concept, JERICO-S3 developed a long-term observation module able to measure a set of common coastal EOVs on the one hand and to integrate different sensor packages adapted to particular fields of study (e.g. Plankton variability, BGC Eutrophication).

The module is adapted to the harsh coastal environment constraints. It is based on the Communication and Storage Front-end (COSTOF2), a platform developed by Ifremer and able to accommodate twelve sensors by providing them controlled power, a common time base, large data storage capacity, communication channels with local or remote users as well as a active anti-biofouling protection.

The platform is able to work in a very low power environment and emphasis is given on the cEGIM ability to process sensor outputs in order to make appropriate decisions, such as triggering sensor acquisition (on or off), tuning acquisition parameters and sending alerts.

The cEGIM has been defined to operate in different coastal environments, like for example underwater (benthic or water column), at the sea surface (buoys) or on jetties and pontoons, either autonomously or wired to an energy/communication facility. It is designed to allow access to a wide library of existing sensor drivers, inherited from coastal and open ocean applications, through procedures for easy integration of new sensors and functions.
The Jerico-S3 intelligent services are aimed at triggering the cEGIM sensors and samplers, adaptively, change their configurations (e.g. data sampling frequency) according to the environmental conditions dynamics and identify and select relevant information into the acquired data. In order to meet this requirements, the cEGIM provides the following services:

- Data time stamping
- Data storage
- Hardware support for the execution of data processing algorithms on the stored data
- Adaptive sensor sampling, based on the results of the data processing algorithms
- Bi-directional communication with the external world either

Details on the cEGIM specification and development can be found in the deliverables:
- D7.1 Prototype of JIIM, with core set of instruments for the measurement of cEOVs
- D7.2 Suite of Sensor Web components for the end-to-end integration of the JIIM into interoperable infrastructures
- D7.3 Test report of JIIM prototype in representative different coastal environments

3.2.Data Processing capabilities within the JERICO-S3 consortium

Underwater observatories are a remarkable case of data-rich environments, as a consequence data science and artificial intelligence approaches are gaining a growing consensus in the marine science and technology community. Autonomous platforms will create an important increase in the production of data (European Marine Board, 2020) and the need for novel and effective technologies, creating a paradigm shift from the traditional vessel-assisted, time-consuming and high-cost sampling surveys (Farcy 2019, Aguzzi 2019).

Data science and artificial intelligence methodologies studied in JERICO-S3 are aimed at improving the observing capabilities of the infrastructure by defining and implementing both a set of intelligent services for the coastal EGIM, and a set of data processing tools accessible or deployed on the JERICO-S3 virtual access e-Infrastructure. The intelligent services may be executed on board the coastal EGIM or in a land laboratory after the data have been transferred from the module. The coastal EGIM is equipped with an embedded processing unit allowing for the onboard execution of specifically developed data processing algorithms that will be the basis for the automated intelligent services.

According to the observed environmental conditions, the proposed services identify and select relevant information from the acquired data and, based on the analysis of such information they may activate sensors and samplers, and adaptively change their configurations (e.g. sampling frequency, resolution). The effectiveness of such services will be demonstrated in an application context mainly dealing with plankton dynamics.

Within the JERICO-S3 consortium many skills for data analysis are available and are summarized in the following table.

Research Activity	**Partner**
Passive Fluorimetry	LISIC, IFREMER-LER/BL, CNRS-LOG
Active Fluorimetry	LISIC, IFREMER-LER/BL, CNRS-LOG
Cytometry data analysis	LISIC, IFREMER-LER/BL, CNRS-LOG
Multivariate time series analysis and change detection	LISIC, IFREMER-LER/BL, CNRS-LOG, CNRS-LOV
Time Series completion	LISIC, IFREMER-LER/BL, CNRS-LOG, HEREON
Multivariate time series analysis and modeling	CNR
Determining regions of representativity in heterogeneous oceanographic datasets	HEREON
Visual data exploration of heterogeneous datasets	HEREON
Plankton classification from images	CNRS-LOV,IFREMER-LER/BL, HEREON
Macro and mega fauna recognition and classification from underwater images	CNR

Passive Fluorimetry: The passive fluorimetry is used for determining phytoplankton groups in relation to their pigment composition and it is a variable of fundamental interest in the JERICO-S3 program. It is studied by spectral fluorimeters at five wavelengths as in the case of the Fluoroprobe bbe Moldaencke and Algae Online Analyser (AOA) system, to provide spectral fingerprints characteristic of four or five microalgae groups: brown algae (Heterokontophyta and Dinophyta), cyanobacteria (cyanobacteria with phycocyanin), green algae (Chlorophyta) and Cryptophyta (Cryptophyta, Rhodophyta, cyanobacteria with phycoerythrin) and finally Prymnesiophyceae (Phaeocystis globosa, in the Eastern English Channel). Biomass of each of four groups (according to real phytoplankton composition of a location) is estimated in an equivalent amount (µg) of Chl a per liter from a mixture assemblage of natural communities.
A Laplacian similarity-based clustering algorithm [Lefebvre et al. 2019] has been defined to detect a variety of profiles for common species or groups and to control outliers. Can be shown that this profile could vary according to specific environmental conditions. AOA and other data sources coupled with an unsupervised spectral clustering method was also used for the discrimination, under near real-time, of 6 to 33 contrasting water masses in the English Channel based on their abiotic and biotic characteristics. AOA and BBE GUI are available on DYMAPHY (www.dymaphy.eu/) website and should be adapted with the new R version 4.0.

Active Fluorimetry: Concerning the active fluorescence, LISIC, IFREMER-LER/BL and CNRS-LOG collaborate on the definition of an R based tool to enable the analysis of measurements by FRRf (Fast Repetition Rate fluorometer) to link the photobiological parameters resulting from these analyses to the multiparameter environmental data of system like Ferry-boxes.

FRRf measurements allow us to obtain photosynthesis-energy relationships (PE), for example every 25 minutes, by a semi-automated method, at large spatial scale during an oceanographic campaign. These PE relationships after analysis provide the photosynthetic parameters of microalgae that allow us to characterize the photo acclimatization state of microalgae and to calculate primary production rates for water columns, with assumptions about the validity of these parameters for a certain spatial area (of the same hydro-biological characteristics) and in the water column (with the vertical variation of light climate, these parameters can indeed vary). The objective of the linkage of physiological and hydrological data is to assess whether and how photobiological and primary production parameters can complement and improve the hydro-biological characterization of coastal seawater bodies.

Cytometry data analysis: RclusTool is a CRAN R package with a GUI (Graphical User Interface) that was deployed by the LISIC team in February 2020. It allows cytometry data to be observed (features, signals and images), and to be (semi-)automatically either clustered or classified (based on the features tables). The R based tool has an open-source license and its main functionalities are:

- data preprocessing (log transformation, arithmetical features combinations, PCA, particles selection, scaling, sampling, etc);
- visualization tools (scatter plots, image and signals views, clustering representations, statistical summary results);
- clustering methods (K-means, Hierarchical Clustering, EM, Spectral Clustering);

- supervised classification methods (K-Nearest Neighbors, Multi-Layer Perceptron, Support Vector Machine, Random Forest);
- semi-supervised methods (Constrained K-Means and Constrained Spectral Clustering);
- batch processing to serially apply a classification method to a set of data files.
- The semi-supervised module of the package relies on Constrained spectral embedding for K-way data clustering [Wacquet et al. 2013a,b]
- A new version is planned at the beginning of 2021, principally to propose a convenient manual clustering.

Multivariate time series analysis and change detection: The uHMM interface [Rousseeuw et al., 2015] and associated packages were published on the CRAN website in April of 2016 (https://cran.r-project.org/web/packages/uHMM/index.html). It aims at identifying general patterns, intermittent or rare events on multivariate time series and proposing a characterization of events and modeling of their dynamics (up to prediction). Two modes - expert and standard are available. In the standard mode, the user has no parameter to tune, and no preprocessing or method to choose. Both event segmentation and clustering could be done in expert mode according to different unsupervised clustering methods (k-means, spectral clustering, hierarchical clustering). From the event segmentation obtained in the clustering step, unsupervised Hidden Markov Models of the time series dynamics are built to improve change detection and to compute dynamic information (state emission probabilities, state transition probabilities).
Many experiments were done to validate the methods on high frequency data from short-duration cruises (1-5 days) to bigger ecosystemic campaigns or buoys data series (IBTS2019 cruises on board RV “Thalassa”, “MAREL Carnot” buoy).
A deep spectral clustering algorithm was proposed to extract a hierarchical view of the dynamics of a process and then applied to phytoplankton biomass. It was explained in [Grassi et al. 2019]. This hierarchy of events offers a general view (e.g. productive vs; non-productive period, dynamics of the spring bloom) and also highlights specific events such as extreme events or rare events. This new algorithm, which combines hierarchical and spectral approaches, gives promising results in the segmentation of both spatial UCI databases and marine time series compared to other approaches. The ability of our method to deal with many kinds of datasets allows a large comparability of results if applied within a broad Integrated Observing System. Beyond scientific knowledge improvements,this comparability is crucial for decision-making about environmental management.

Time series completion: Much attention has been paid to signal pre-processing to improve clustering but also prediction. The existence of missing data is a common fact in real applications which can significantly affect the data analysis, especially when the percentage of missing data is high and the length of the gaps is large. In order to overcome this problem, many methods have been proposed in the literature. Several algorithms of completion were tested from simple imputation based on the last observation to more complex ones, such as multiple imputation or machine learning techniques as, for example, GAN and RandomForest. Two new algorithms based on window retrieval and elastic similarities were proposed for multivariate signals (R CRAN packages DTWUMI and FSMUMI)

- https://cran.r-project.org/web/packages/DTWUMI/index.html)
- https://cran.r-project.org/web/packages/FSMUMI/index.html

Now, LISIC Team with IFREMER-LER/BL are going to build a system able to choose the best algorithm that suits from extracted signal features of the dataset and its type of missingness.
Continuous data collection at the PSSs, through long-term instrument deployments such as the one at the English Channel and North Sea PSS, contributes to the accumulation of highly resolved visual and biochemical data time series. Although data collection efforts spanning only a few years in length may be considered too short to qualify as true long-term time-series, the life cycle of the planktonic organisms of interest is short enough to be covered sufficiently by this time span.

Multivariate time series analysis and modeling: Ecosystems dynamics are an integrated response of the ecosystem's biological components (species/groups) to drivers, which act independently, synergistically or even antagonistically. These drivers are defined as any natural or human-induced factor that directly or indirectly causes a change in an ecosystem or population.
The multivariate analysis approach proposed in [Papworth2016] is a data-driven modeling methodology built on a symbolic regression approach based on genetic programming (GP). It is aimed at discovering the most relevant ecosystem drivers, starting from a large set of candidates environmental parameters and aimed at building a model based on the identified drivers. The obtained model consists of one or more explicit mathematical expressions approximating a given environmental variable (target variable) as function of other environmental variables (drivers). All the involved variables are expressed as time series. Differently from the traditional regression methodologies, the proposed symbolic regression based on GP does not need to assume hypotheses on the mathematical form of the approximating functions and also does not need strong a-priori knowledge on the specific application domain. This data-driven modeling methodology can be used for combining environmental, physical, biological and biogeochemical time series into an explicit explanatory model for capturing and understanding the temporal dynamics of the whole ecosystem or of single species or populations. The obtained model can also be used for predicting the target variable trend as a function of the identified drivers

Determining regions of representativity in heterogeneous oceanographic datasets: Existing sensor networks supply marine data centers with a large amount of autonomous measurements of oceanographic parameters, such as temperature, salinity, chlorophyll and oxygen saturation. Clustering algorithms can help identify patterns in the data, determine regions of similar conditions and possibly help to strategically expand the measurement network.
In a first step, we identified regions of representativity around Helgoland [Wichert, Brix. 2020] using an unsupervised machine learning algorithm. The regions are determined by clustering a dataset consisting of multivariate time series from the measurement network around Heligoland as well as complementary model time series from that area. Further efforts will focus on analyzing the temporal stability of the areas of representativity and the influence of individual parameters on the regional pattern. Results will be applied to strategically place sensors in under-represented areas of the measurement network.

Visual data exploration of heterogeneous datasets: The Flood Explorer, developed in the Digital Earth project [https://www.digitalearth-hgf.de/] consists of several interactive workflows that focus on different stages of a flood event. The "River Plume Workflow" [https://www.digitalearth-hgf.de/workflows] is developed at HEREON. It focuses on the later stages of a riverine flood event, when river water flows into the Sea. The River Plume Workflow provides an interactive user-interface for detecting the river plume in FerryBox trajectories in the North Sea and finding observational data to quantify the influence of the floodwater's influx into the marine environment. Users are provided with functions to find the river plume in the FerryBox data and perform a preliminary analysis to compare its composition to the surrounding waterbodies. Due to the use of additional model data, information on the river plume's spatio-temporal propagation can be gathered and used to identify relevant observational datasets for further analysis. The River Plume Workflow helps scientists to analyze past flood events across a range of different observational datasets and offers support for near-future campaign planning.

3.3.Intelligent services and data analysis algorithms for the cEGIM

The data science methodologies studied in the task T7.4 are aimed at defining and implementing both a set of intelligent services for the coastal EGIM module (cEGIM) and a set of data processing tools to be exposed within the virtual access e-Infrastructure of J-S3. According to the requirements defined in D7.1 "Prototype of JIIM, with a core set of instruments for the measurement of cEOVs." The cEGIM consists of an interoperable module for measuring different underwater environmental variables. The module is capable of hosting different sensors and samplers, while respecting the JERICO-RI data flow and interoperability principles and being able to host the intelligent services planned within the Jerico-S3 objectives.

Within the WP1 activities, several Key Scientific Challenges (KSC) based on relevant scientific questions (case studies) have been proposed by the JERICO community. Among the proposed KSC, the WP1 and WP7 leaders selected those focused on phytoplankton dynamics analysis and, based on these, defined an application scenario for validating the cEGIM effectiveness. Based on the proposed scenario, the WP7 participants selected the variables to be investigated, the sensors to be installed on the cEGIM and the corresponding sensor acquisition frequency. This information has been used for setting up the cEGIM demonstration and for defining the cEGIM intelligent services and data processing algorithms implemented in the T7.4 research activities.

The intelligent services have been defined as a set of stand-alone software tools that should be automatically executed on the data collected by the cEGIM and aimed at:

- Switching on/off the cEGIM sensors;
- Changing the cEGIM sensor configuration (e.g. acquisition frequency);
- Identifying relevant events from the acquired data (e.g. time series change points);
- Transferring the acquired data to the e-JERICO infrastructure (JERICO-CORE).

The intelligent services are executed on board the cEGIM or in the lab, after the data are transferred from the cEGIM. In the latter case, the cEGIM provides a double-link connection with the land station in order to transfer the acquired data and to receive command for the hosted sensors. Independently by the communication capabilities, the cEGIM also provides an API allowing the software tools to interact with the hosted sensors.

3.4. Application scenario for the phytoplankton dynamics analysis

The objective of the application scenario defined for validating the cEGIM is the investigation of the primary and secondary springtime phytoplankton blooms at the MAREL SMILE Buoy[1], in the English Channel, near the mouth of the Seine river. The investigation of the secondary bloom is particularly relevant for studying the mechanisms triggering the bloom of harmful Algae.
The sensors installed on the cEGIM and used for the bloom analysis are listed in the following Table:

Manufacturer	Reference	Main measurements	Output format
NKE	MP6	Depth, Temperature, Conductivity, Dissolved Oxygen, Turbidity	Modbus
BBE moldaenke	Fluoroprobe	chlorophyll with algae class determination - green algae, blue-green algae/cyanobacteria, diatoms/dinoflagellates and cryptophytes	ASCII
SATLANTIC	MBARI-ISUS	Nitrate concentration	ASCII
Hydroptic / LOV	UVP6	Micro zooplankton	images

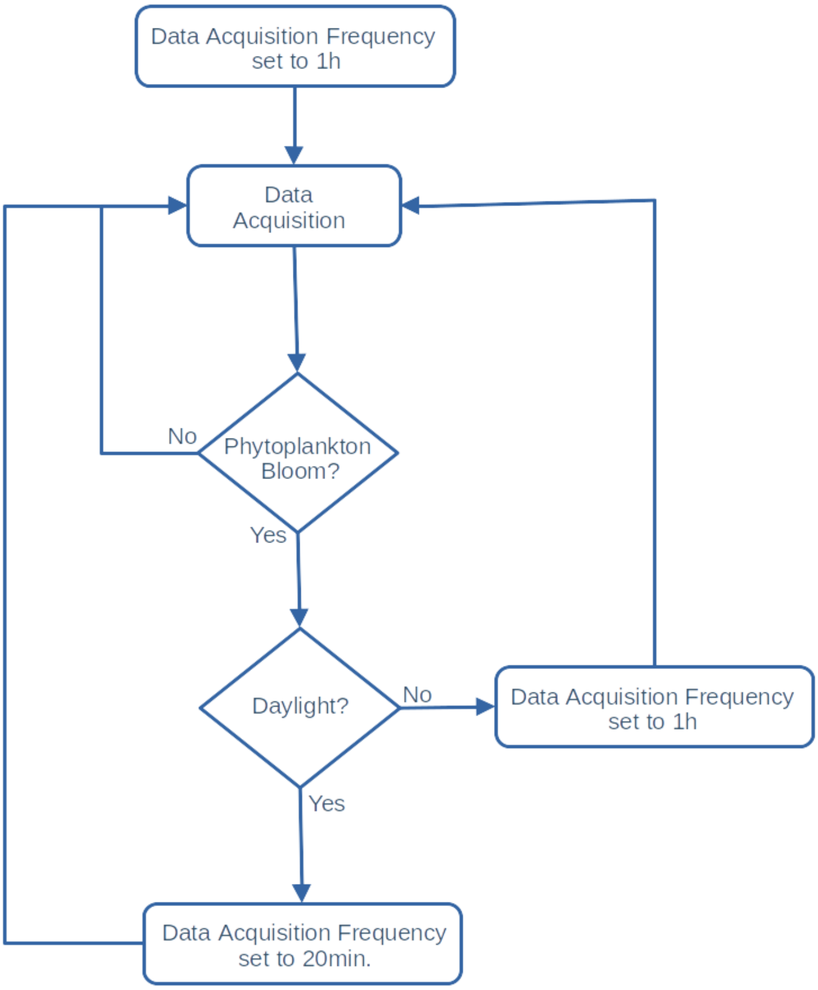
In order to automatically analyze the phytoplankton bloom, the Start-Bloom Triggering Event, the End-Bloom Triggering Event and the cEGIM Adaptive Behavior have been defined as in the following:

Start-Bloom Triggering Event: sudden decrease in nutrients combined with a sudden increase in fluorescence. This event represents the beginning of the bloom.

End-Bloom Triggering Event: increment of nutrients combined with the decrement of fluorescence. This event represents the end of the bloom.

cEGIM Adaptive Behavior: at the beginning of the monitoring session, all the sensors onboard the cEGIM acquire data every hour, when the Start-Bloom Triggering Event is detected all the sensors increase the acquisition frequency to one measurement every 20 minutes. When the End-Bloom Triggering Event is detected the acquisition frequency of all the sensors is set back to one measurement every hour. The following schema summarize the behavior of the cEGIM according to the defined application scenario:

[1] https://www.seanoe.org/data/00425/53689/



The algorithm for detecting the phytoplankton bloom is based on the research activity present at the following link and discussed in section 3.2:

https://mawenzi.univ-littoral.fr/

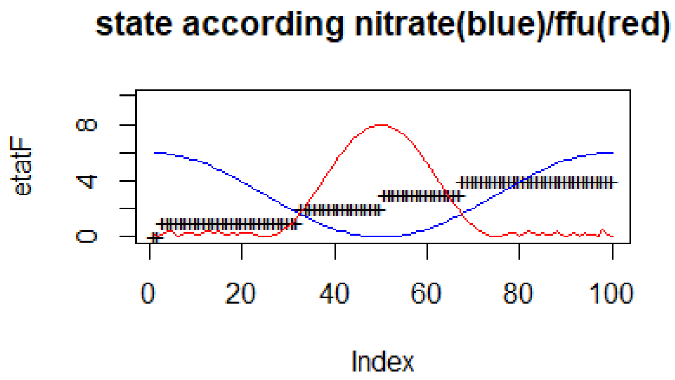
Such tools are ready to be inserted in the J-CORE catalog.

The algorithm implementation is available on the GitLab platform at the following link:

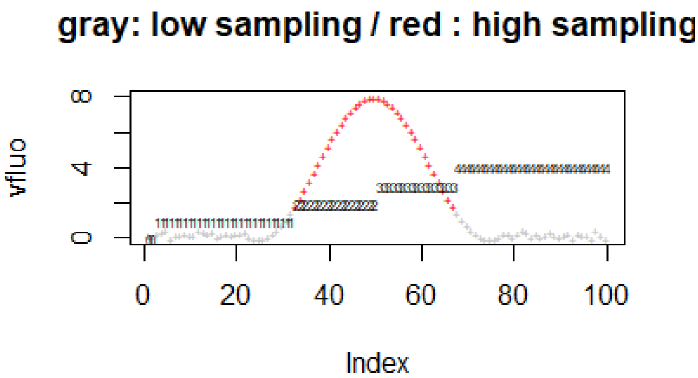
https://gitlab.com/epoisson/cegim

The algorithm takes as input the equivalent-chlorophyll-*a concentration* and the nitrate concentration acquired by the BBE Moldaenke Fluoroprobe and the SATLANTIC MBARI-ISUS sensors, respectively and, returns as output a flag value about the bloom status. In particular If the output flag is equal to the value +1, the cEGIM have to increase sensors' sampling frequency; if the output flag is equal to the value -1, then the cEGIM has to decrease the sensors' sampling frequency; finally if the flag is equal to the value 0, then the sensors' sampling frequency remains unchanged.

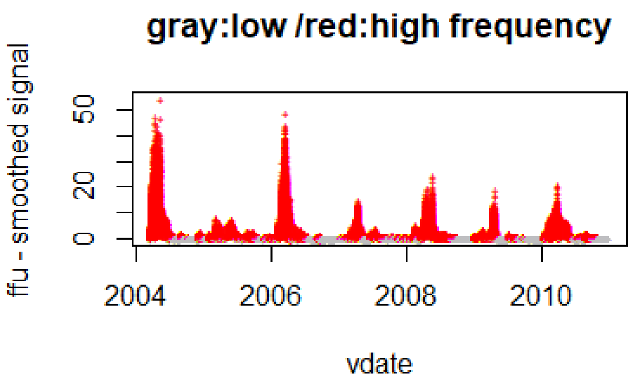
The following figure shows the behavior of the bloom detection algorithm on the historical data available at the MAREL SMILE buoy. The panel a) shows the nitrate values (blue line) and the fluorescence values (red line) during a phytoplankton bloom. The panel b) shows an example of fluorescence sampling at different frequencies according to the output of the bloom detection algorithm. Gray dots represent measurements at low sampling frequency (not detected bloom), while red dots represent measurement at high sampling frequency (detected bloom). Panel c) show the algorithm behavior on the historical data at the MAREL SMILE Buoy in the period 2004-2011. Again, the red dots represent the detected blooms, while the gray dots represent not bloom periods.



a)



b)



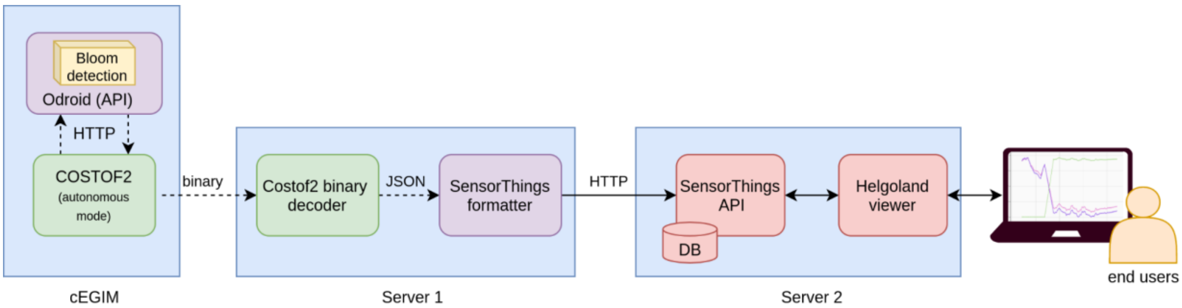
c)

After the data has been acquired by the cEGIM it is transferred to the e-JERICO infrastructure. Data from COSTOF2to shore station is transmitted using an efficient binary protocol. This protocol is envisioned to save bandwidth (and energy) in constrained scenarios such as long deployments powered with batteries using acoustic communications.

Binary data is received and decoded at the shore station server, where it is formatted and sent via HTTP to an OGC [SensorThings API](https://docs.ogc.org/is/18-088/18-088.html) (https://docs.ogc.org/is/18-088/18-088.html). This service provides an open, geospatial-enabled and unified way to interconnect devices, data, and applications over the Web (https://github.com/52North/sensorweb-server-sta). This service is in charge of storing all data and associated metadata, providing a coherent and robust access and storage mechanism for geospatial data in real-time.

Further services can access data and metadata by using the SensorThings API endpoints. As an example of a visualization tool a Helgoland data viewer has been deployed (https://github.com/52North/helgoland).

The following figure illustrates the data pipeline, from the cEGIM to the end users.



4. OUTREACH, DISSEMINATION AND COMMUNICATION ACTIVITIES

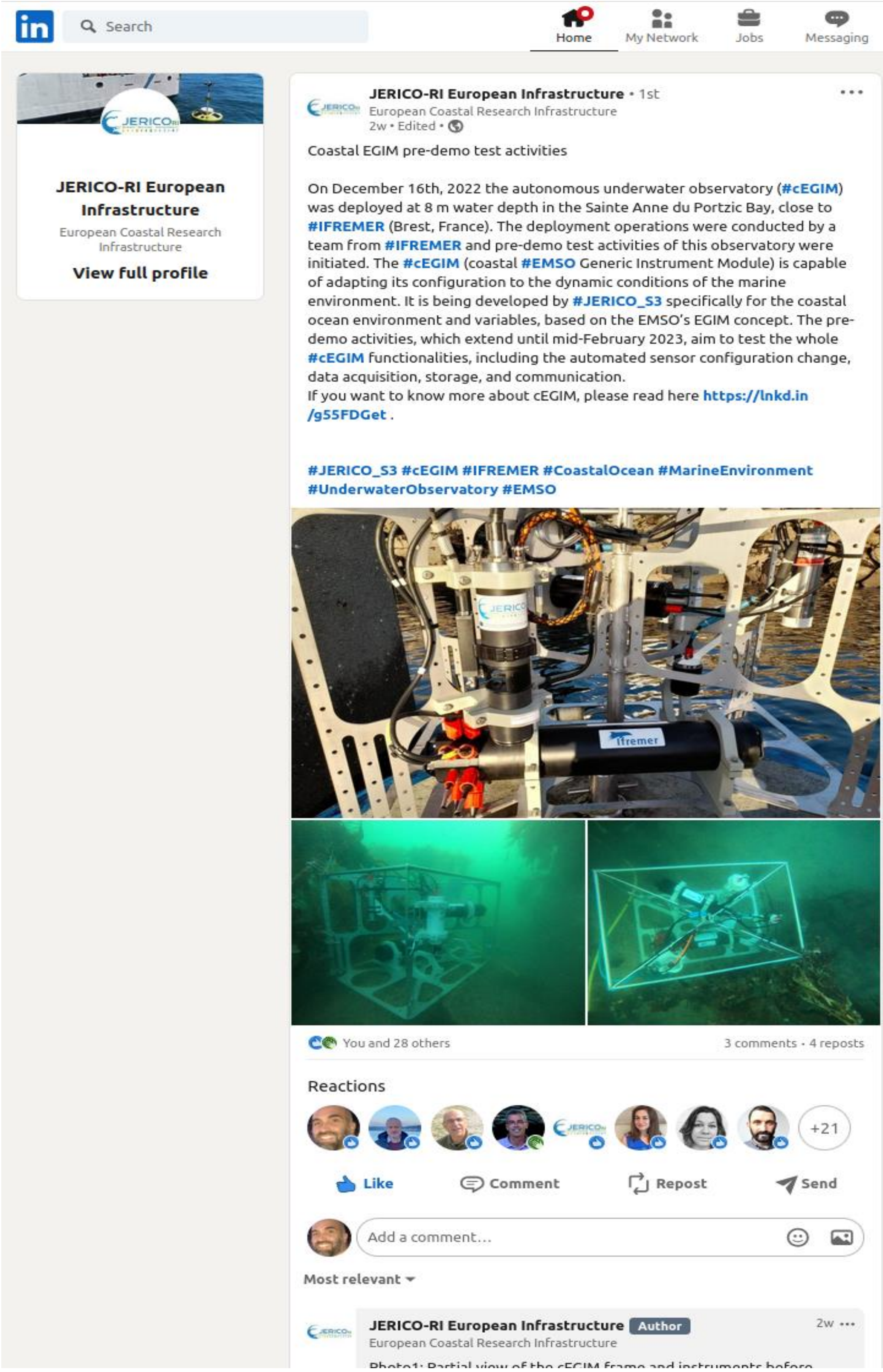
Within the WP7 activities the following contribution has been published in the Proceedings of the 9th EuroGOOS International Conference 'Advances in Operational Oceanography: Expanding Europe's Observing and Forecasting Capacity':

https://eurogoos.eu/download/eurogoos-conference-2021-conference-proceedings/

Delory Eric, Marini Simone, Blandin Jerome, Boccadoro Catherine, Durand Miguel, Cianca Andres, Tintore Joaquin, Pearlman Jay, Charcos Miguel, Alcalde Miguel Angel, Fernandez Juan Gabriel, Delauney Laurent (2021). **JERICO-S3 Integrated Innovative Technologies for Coastal Monitoring**. Fernandez Vicente, Lara-Lopez Ana, Eparkhina Dina, Cocquempot Lucie, Lochet Corine, Lips Inga (eds.) (2021). Proceedings of the 9th EuroGOOS International Conference 'Advances in Operational Oceanography: Expanding Europe's Observing and Forecasting Capacity'. 3 – 5 May 2021. EuroGOOS. Brussels, Belgium. 2021. pp.186-192. https://archimer.ifremer.fr/doc/00804/91591/

Moreover, at the beginning of the test activities of the cEGIM a post on the first deployment of the cEGIM has been published on Linkedin:

https://www.linkedin.com/feed/update/urn:li:activity:7018205238641778688/



New dissemination activities are planned after the end of the demonstration activities within the Task T7.6 when new data will be gathered by the cEGIM and the intelligent services will be assessed and validated.

5.CONCLUSIONS

The coastal marine environment is complex. Yet many resources for human well being come from the coast as well as risks for coastal communities. JERICO-S3 is focused on providing a foundation for applications and informed policy and decision making. Advances in observation technology and access to data and information are key to provide a comprehensive and trusted resource. The intelligent services designed within JERICO-S3 address the biological dimension from the point of view of autonomous sensors and adaptive sampling, and the integration of all resources to advance knowledge under a common environment, connected with external data brokers, services, and users.

The present deliverable shows the current development of the intelligent services the cEGIM is equipped with. Such services have been designed according to the interest of the JERICO-S3 community for the phytoplankton analysis and are based on the data acquired by the sensors selected to be installed onboard of the cEGIM for the demonstration activity planned in the English Channel super site. This deliverables is also highlighting recent methodological numerical development based on Machine Learning which aim at optimizing the pre-processing and processing steps when handling when considering data as complex as high frequency data. Such tools are accessible or deployed on the JERICO-S3 virtual access e-Infrastructure.

Such services consist of an algal bloom detection algorithm, the algorithm implementing the dynamic behavior of the cEGIM, the driver implementing the sensors' adaptation and the communication and visualization pipeline involving the acquired data.

Moreover, the present deliverable also captures the data analysis skills within the JERICO-S3 consortium that poses the bases for further implementations of improved intelligent services on a wider set of potential application contexts.

6.ANNEXES AND REFERENCES

[Grassi et al. 2019] Grassi, K.; Poisson-Caillault, É.; Bigand, A.; Lefebvre, A. Comparative Study of Clustering Approaches Applied to Spatial or Temporal Pattern Discovery. J. Mar. Sci. Eng. 2020, 8, 713. https://doi.org/10.3390/jmse8090713

[Lefebvre et al. 2019] Lefebvre A., Caillault-Poisson E., 2019. High resolution overview of phytoplankton spectral groups and hydrological conditions in the eastern English Channel using unsupervised clustering. Marine Ecology Progress Series. Volume 608, page 73-92, https://doi.org/10.3354/meps12781

[Papworth et al. 2016] Papworth, D.J., Marini, S., Conversi, A. A novel, unbiased analysis approach for investigating population dynamics: A Case Study on Calanus finmarchicus and Its Decline in the North Sea (2016) PLoS ONE, 11 (7), art. no. e0158230, .

[Rousseeuw et al, 2015] Rousseeuw, K.; Poisson Caillault, É.; Lefebvre, A.; Hamad, D. "Hybrid Hidden Markov Model for Marine Environment Monitoring". In : IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8.1 (jan. 2015). DOI : 10.1109/JSTARS.2014.2341219.

[Wacquet et al. 2013a] Wacquet G, Caillault EP, Hamad D, Hebert P (2013). "Constrained spectral embedding for K-way data clustering." Pattern Recognition Letters, 34(9), 1009–1017. doi:

10.1016/j.patrec.2013.02.003, https://hal.archives-ouvertes.fr/hal-01536663.

[Wacquet et al. 2013b] Wacquet G, Caillault EP, Hebert P (2013). "Semi-supervised K-Way Spectral Clustering with Determination of Number of Clusters." In Madani, Kurosh, Dourado, Antonio, Rosa, Agostinho, Filipe, Joaquim (eds.), Computational Intelligence: Revised and Selected Papers of the International Joint Conference, IJCCI 2011, Paris, France, October 24-26, 2011, chapter Semi-supervised K-Way Spectral Clustering with Determination of Number of Clusters, 317–332. Springer Berlin Heidelberg, Berlin, Heidelberg. ISBN 978-3-642-35638-4, doi: 10.1007/978-3-642-35638-4_21.

[Wichert, Brix. 2020] Wichert, V. and Brix, H.: Augmenting the sensor network around Helgoland using unsupervised machine learning methods, EGU General Assembly 2020, Online, 4–8 May 2020, EGU2020-15212, https://doi.org/10.5194/egusphere-egu2020-15212, 2020.